Editors' note: *This series addresses topics that affect epidemiologists across a range of specialties. Commentaries are first invited as talks at symposia organized by the Editors. This paper was originally presented at the 2007 Society for Epidemiologic Research Annual Meeting.*

Emerging Technology in Molecular Epidemiology

What Epidemiologists Need to Know

Frederica P. Perera and Julie B. Herbstman

Abstract: Epidemiology has evolved with the development of new molecular technologies that refine the way we investigate the relationships between exposure and disease. While these novel tools open new opportunities to delve into the mechanisms of the molecular epidemiologic continuum, they also come with the challenge of ensuring their meaningful application in epidemiologic investigations. To train successful molecular epidemiologists in the postgenome/epigenome era, we can look to the "lessons learned" when epidemiologists first integrated molecular biomarkers into traditional epidemiologic designs. These examples show how interdisciplinary training in molecular epidemiology programs can help ensure that new technologies are used effectively to understand the mechanisms driving exposure-disease relationships.

(*Epidemiology* 2008;19: 350–352)

In the last 25 years, the practice of epidemiology has evolved with the development of new molecular technologies that have allowed us to refine the way we investigate the relationships between exposure and disease. Advances in molecular biology have increased in a seemingly exponential fashion following the identification of the DNA double helix by Watson and Crick in 1953. These innovations have provided the basis for many of the tools used in modern epidemiology, leading to the formal introduction of the concept of molecular epidemiology in 1982.[1] Molecular epidemiologic tools have enabled us to explore the mechanistic pathways that underlie observed exposure-disease relationships that were formerly hidden in a "black box." Initial research focused on various forms of direct genetic damage as relevant biomarkers.

With the completion of the Human Genome Project[2] and the initiation of the Human Epigenome Project,[3] molecular tools have expanded further, providing modern-day molecular epidemiologists with powerful new laboratory-based techniques that include epigenetic and "-omics" technologies (genomics, proteomics, metabonomics, etc.). While these novel tools provide new opportunities to delve deeper into the components of the molecular epidemiologic continuum, they also come with the challenge of ensuring their meaningful application in modern epidemiologic investigations.

In many ways, learning to incorporate the emerging technologies of today is very similar to the way in which epidemiologists added tools to their toolboxes when they first integrated molecular biomarkers into traditional epidemiological designs. To train successful molecular epidemiologists in the postgenome/epigenome era, we can look to the "lessons learned" from that transition to guide in integrating today's emerging technologies. Using the example of the successful incorporation of the biologic markers of DNA damage related to polycyclic aromatic hydrocarbons (PAH-DNA adducts), we highlight the critical importance of biomarker validation as well as the continued value of basic epidemiologic concepts and framework.

PAH-DNA Adduct Measurements in Epidemiology: Lessons Learned

PAHs are potent carcinogens found in tobacco smoke and other environmental mixtures.[4] To study health effects associated with PAH exposure, PAHs could be measured in the air or estimated using a questionnaire. Miller and Miller[5] established experimentally that PAHs can bind covalently with DNA, forming PAH-DNA adducts, and that DNA adduct formation was a causal carcinogenic mechanism in laboratory animals. The rapid quantification of PAH-DNA adducts was made possible with the development and application of an enzyme-linked immunosorbent assay.[6] Using this new technology, in 1982 PAH-DNA adducts were detected and measured in a human population in vivo—providing an opportunity for epidemiologists to use this biomarker

Submitted 4 September 2007; accepted 25 October 2007.
From the Department of Environmental Health Sciences, Columbia Center for Children's Environmental Health, Columbia Mailman School of Public Health, New York, NY.
NIEHS and EPA grants: 1 RO1 ES008977, 5 PO1 ES009600, 5 RO1 ES10165; 5 RO1 ES11158; 5 RO1 ES012468, 2 EPA Star grants: R-82860901 & R-827027, RD-832141, P30ES006096; R21ES013071; R21ES013071 and Private foundations.
Editors' note: *Related articles appear on pages 345, 347, and 353.*
Correspondence: Julie B. Herbstman, 100 Haven Avenue, #25F, New York, NY 10032. E-mail: jh2678@columbia.edu.

ISSN: 1044-3983/08/1902-0350
DOI: 10.1097/EDE.0b013e318162a920

in their investigations of PAH-related disease.[7] The possibility of incorporating a biologic marker that reflected the biologically effective dose of PAH marked a substantial improvement from the more traditional methods of assessing PAH exposure for epidemiologic research.

However, before investigators could use PAH-DNA adducts as biomarkers in epidemiologic studies, molecular epidemiologists carefully examined the characteristics and validity of the laboratory methodology, including the sensitivity, specificity, minimum quantify of DNA required for quantification, limit of detection, and factors that might compromise the accuracy of the measurement. Once the assay had been characterized, a series of validation studies were undertaken before the biomarker was used in large-scale epidemiologic investigations. Although it was clear that the PAH-DNA adducts could be measured in human blood, could levels of adducts distinguish between exposed and unexposed populations? Between cancer cases and controls? Were adducts measured in peripheral white blood cells good surrogates for adducts in target tissues?[8–10] How much of the variability in DNA adduct measurements were due to between-person, within-person, or laboratory variability?[11] It is clear that before PAH-DNA adducts were planned for use as biomarkers in full-scale epidemiologic studies, there were many preliminary validation studies that were undertaken. Each of these steps involves proficiency in academic disciplines in addition to basic epidemiologic training, including molecular biology, toxicology, laboratory science, and biostatistics.

While proficiency in these areas is important for designing and understanding the implications of validation studies undertaken prior to the design and initiation of a full-scale epidemiologic study, they are also critical for trouble-shooting, as the full-scale study progresses. For example, during a longitudinal study in which PAH-DNA adducts were used as biomarkers of biologically effective dose, the laboratory methodology improved and the low-dose sensitivity of the assay increased. From the laboratory perspective, it made sense to use the improved assay. However, a well-trained molecular epidemiologist would identify the potential fatal flaw in changing methodologies "midstream."

This exemplifies the way in which the epidemiologist needs to be involved in and understand all aspects of the study, even those moving outside of traditional epidemiologic training.

Using Lessons Learned to Guide Incorporation of New Technology

Today's modern epidemiologists face similar challenges as they assess the potential for using new technologies in epidemiologic studies.[12] For example, epigenetic modifications, described as heritable changes in the genome that do not involve alterations in nucleotide sequences, have emerged as a promising explanation for the observed variation between genotype and gene expression.[13] Epigenetic modifications take many forms, including aberrant gene promoter methylation, which may impact the ways genes under the control of this promoter region may be expressed. To evaluate this mechanism, many laboratory methods have been developed to assess various aspect of DNA methylation, including global methylation (describing a participant's overall methylation fingerprint) and gene-specific methylation (describing the methylation of DNA regions that control specific gene expression). The modern epidemiologist must answer a series of questions before entertaining the idea of incorporating these markers in large-scale studies.

First, the epidemiologist is required to select a laboratory method. As with PAH-DNA adduct measurements, there are advantages and disadvantages to each of the laboratory techniques designed to quantify DNA methylation (as outlined by Ho and Tang[14]). The selection of the method for use in an epidemiologic study is dependent on many factors, including those that are directly related to laboratory characteristics (the assay's sensitivity, specificity, reproducibility) as well as other logistical characteristics (including the amount of sample required, the cost per run, and the availability of equipment and trained technicians). These issues are similar to considerations molecular epidemiologists faced when incorporating PAH-DNA adducts.

One of the biggest differences between the first uses of PAH-DNA adducts in epidemiology and the incorporation of today's emerging technology is the quantity of data that is generated. The field of bioinformatics developed as a result of the data generated from new laboratory techniques, and it will be crucial to have as a coinvestigator a biostatistician or another individual well-trained in interpretation of this data. In addition to necessitating additional expertise, the generation of this type of data also shifts the validation paradigm to one of discovery as well as hypothesis-testing. Initially discovery-oriented approaches must be used to sort through the vast quantity of information that is generated from these new genome (or proteome or metabonome)-wide approaches to determine relevant patterns and biomarkers for use in hypothesis-testing.

Epidemiologists must develop a deep enough understanding of the principles of these disciplines to evaluate when to use and when NOT to use biomarkers generated from these new technologies. Although tempting to incorporate new markers because they are "new and exciting," epidemiologists need to develop the skills to know when the biomarkers have been sufficiently validated so that their interpretation is meaningful.

In a sense, a molecular epidemiologist operates as the conductor of a scientific orchestra of players, including laboratory scientists and technicians, biostatisticians, as well as experts in bioinformatics. Just as it is impractical to expect an orchestra conductor to be able to play every instrument, it is

not reasonable to expect a molecular epidemiologist to become an expert in each of the disciplines contributing to modern molecular epidemiologic research. However, while an orchestra conductor is not required to play every musical instrument, she must understand the sound each instrument makes to coordinate them into a symphony. Similarly, in addition to a mastery of epidemiology, which remains the basis of modern molecular epidemiologic research, a molecular epidemiologist's job is multidisciplinary, requiring proficiency in the fundamentals of each of the disciplines contributing to the research. Incorporating this interdisciplinary training in molecular epidemiology programs will ensure that new technologies can be used effectively to enhance the ability of epidemiologists to draw conclusions about mechanisms driving exposure-disease relationships. Then disease prevention will be an attainable goal.

ABOUT THE AUTHORS

FREDERICA PERERA is Professor in the Department of Environmental Health Sciences at Columbia University, and Director of the Columbia Center for Children's Environmental Health. Her research is on environmental causes of cancer and developmental disorders, including effects of ambient air pollution and pesticides on child health. JULIE HERBSTMAN is an environmental epidemiologist and research fellow in the same department, and directs the World Trade Center Pregnancy Study. Her research focus is the effects of prenatal exposures on child growth and development.

REFERENCES

1. Perera FP, Weinstein IB. Molecular epidemiology and carcinogen-DNA adduct detection: new approaches to studies of human cancer causation. *J Chronic Dis*. 1982;35:581–600.
2. Genomics and Its Impact on Science and Society: A 2003 Primer. In: Human Genome Project USDoE, ed; 2003.
3. Bradbury J. Human epigenome project–up and running. *PLoS Biology*. 2003;1:e82.
4. Bostrom CE, Gerde P, Hanberg A, et al. Cancer risk assessment, indicators, and guidelines for polycyclic aromatic hydrocarbons in the ambient air. *Environ Health Perspect*. 2002;110(suppl 3):451–488.
5. Miller EM, Miller JA. Mechanisms of chemical carcinogenesis. *Cancer Res*. 1981;47:1055–1064.
6. Santella RM. Application of new techniques for the detection of carcinogen adducts to human population monitoring. *Mutat Res*. 1988;205: 271–282.
7. Perera FP, Poirier MC, Yuspa SH, et al. A pilot project in molecular cancer epidemiology: determination of benzo[a]pyrene–DNA adducts in animal and human tissues by immunoassays. *Carcinogenesis*. 1982;3: 1405–1410.
8. Tang D, Santella RM, Blackwood AM, et al. A molecular epidemiological case-control study of lung cancer. *Cancer Epidemiol Biomarkers Prev*. 1995;4:341–346.
9. Veglia F, Matullo G, Vineis P. Bulky DNA adducts and risk of cancer: a meta-analysis. *Cancer Epidemiol Biomarkers Prev*. 2003;12:157–160.
10. Wiencke JK, Thurston SW, Kelsey KT, et al. Early age at smoking initiation and tobacco carcinogen DNA damage in the lung. *J Natl Cancer Inst*. 1999;91:614–619.
11. Dickey C, Santella RM, Hattis D, et al. Variability in PAH-DNA adduct measurements in peripheral mononuclear cells: implications for quantitative cancer risk assessment. *Risk Anal*. 1997;17:649–656.
12. Vineis P, Perera F. Molecular epidemiology and biomarkers in etiologic cancer research: the new in light of the old. *Cancer Epidemiol Biomarkers Prev*. 2007;16:1954–1965.
13. Feinberg AP, Tycko B. The history of cancer epigenetics. *Nat Rev Cancer*. 2004;4:143–153.
14. Ho SM, Tang WY. Techniques used in studies of epigenome dysregulation due to aberrant DNA methylation: an emphasis on fetal-based adult diseases. *Reprod Toxicol*. 2007;23:267–282.